AD 721 235

A COMPARISON OF DEMAND FORECASTING TECHNIQUES

John Ainsworth Coventry

Naval Postgraduate School
Monterey, California

March 1971

United States Naval Postgraduate School

THESIS

A COMPARISON OF DEMAND FORECASTING TECHNIQUES

by

John Ainsworth Coventry

Thesis Advisor: P. W. Zehna

March 1971

Approved for public release; distribution unlimited.

77

Best
Available
Copy

ACCESSION for	
CFSTI	WHITE SECTION ☐
DDC	BUFF SECTION ☐
UNAN.	☐
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
DIST.	AVAIL. and/or SPECIAL
A	

A Comparison of Demand Forecasting Techniques

by

John Ainsworth Coventry
Captain, United States Army, MSC
B.A., University of Chattanooga, 1965

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1971

Author ______________________

Approved by: ______________________ Thesis Advisor

______________________ Chairman, Department of Operations Analysis

______________________ Academic Dean

ABSTRACT

For periodic review inventory models with stochastic demand, the idea of stock-out risk is defined, from which the importance of accurate prediction of demand is deduced. Methods of demand model parameter estimation are investigated and several methods compared on the basis of theoretical soundness, ease of application, and accuracy of estimates based upon the results of extensive computer simulation. The theoretical development of maximum likelihood and exponential smoothing estimators as applied to prediction is presented along with the development of a new Bayesian approach to the problem of demand forecasting.

TABLE OF CONTENTS

I. INTRODUCTION

There is a great deal of interest today in business and industry, as well as government activities, in inventory control. The current literature of operations research contains considerable ideas, methods, mathematical models and algorithms addressing this particular subject area. One of the most difficult problems involved in the application of any mathematical model to actual inventory analysis or control problems is the estimation of the parameters used in the formulas of the models and algorithms. To resolve this problem operations researchers have concerned themselves with the development of accurate estimation techniques and demand forecasting methods. One of the techniques which has received a good deal of attention in the current literature is exponential smoothing, a method of estimation and forecasting developed by Brown [Ref. 2]. The basic idea of the smoothing approach is to use a weighted moving average of demand data in such a way as to exponentially discount the oldest data in the demand sequence and place most of the weight for an estimate on the most recent data available. This exponential discounting of past demands is controlled by the selection of a smoothing constant which determines the relative value of the weight assigned to the data in the estimate. The exponential smoothing technique has been questioned on theoretical grounds by Zehna [Ref. 10] and others. It has been suggested that the classical maximum likelihood methods will produce better results in a

number of demand models and some simulation results are available which indicate that this is the case [Refs. 6 and 11].

The purpose of this thesis is to examine the problem of forecasting demand and explore several alternative methods which should be considered as possible forecasting techniques for particular demand models. A simulation analysis of maximum likelihood, smoothing, and Bayesian estimation methods was conducted in order to ascertain important or noteworthy properties of the techniques when applied to normally distributed random demand in the constant mean and linear demand models.

Throughout this report the idea of a risk involved in inventory operations plays a central role. The demand models studied are assumed to be applied to an inventory system in which stock is controlled by means of a periodic review of inventory levels. That is, at certain specified periods of time, monthly, quarterly, etc., the demand for items is determined along with the current inventory balance. These data are used to requisition replacement stock to insure that the system does not run out of stock too often. If the demand were deterministic there would be no problem in deciding how much to order. However, the demand in this case is assumed random and predictions of the demand in future periods must be made in order to set reorder levels. It is assumed that resupply is instantaneous so that, if it can be determined at the end of a period that within a specified probability the demand in the next period will be less than a certain quantity, then the

reorder level is that expected upper limit on the next period's demand. The reorder quantity is the reorder level less the amount of stock on hand. By deciding what probability of running out of stock, called stock-out risk, is acceptable, the decision maker sets a risk level that determines the ordering policy. Consequently estimating parameters and predicting demand in general is a central issue in the determination of reasonable reorder policies.

In Section II. the theoretical development of the maximum likelihood estimation technique is presented along with a simplified version which was used by Zehna [Ref. 11] for comparison with exponential smoothing. Section III. is a summary of the theoretical basis for exponential smoothing developed by Brown [Ref. 2]. In addition, at the end of Section III smoothing is applied to a case not previously considered by Brown. Section IV presents a comparison of maximum likelihood and smoothing with consideration given to theoretical basis, ease of application, and simulation results. In Section V a Bayesian procedure for demand estimation and prediction is developed for the case when the variance of the demand distribution is known. The resulting Bayes estimation method is compared to exponential smoothing on the basis of computer simulation.

Since simulation was used throughout this study a special section on this technique is presented at the beginning of Section IV. It should be noted that in previous cases studied by Zehna [Ref. 11] and Ornek [Ref. 6], simulation was also used as one method of comparing MLE and smoothing. There

is a basic difference in the approach in those papers and the approach taken in this study. Their procedure was to generate random demand over a long period of time, 1000 periods, making estimates with the various methods at each period, and determining the sample risk (probability of running out of stock) over the entire time of operation. The approach taken here, however, realizing that in many applications only a limited amount of data may be available or used as a basis for estimation, is to generate demand data, make estimates, and determine the risks involved at various points in time, say 5 periods, 20 periods, etc., over a large number of replications of the experiment. It is felt that the results of this simulation method should prove useful in determining which estimating techniques perform well when estimates are based upon a fairly limited amount of data.

II. MAXIMUM LIKELIHOOD ESTIMATION

A classical approach to forecasting or predicting the value of a random variable based upon past observations is to develop a functional relationship between the observed values of the random variable and some control variable using regression analysis. Applied to the present circumstances one may view the random variable Y as demand and the control variable x as time. Demand is observed at various times $x_1, x_2, \cdots, x_n$ and then inferences are made about the dependence of Y on x.

In this section the results of the General Linear Hypothesis are examined as they apply to forecasting demand when the mean demand is a linear function of time, that is, $E(Y_i) = a + bx_i$.

A. GENERAL LINEAR MODEL

1. Estimate of the Coefficients

$$\text{If } \underline{Y} = \underline{A\beta} + \underline{\varepsilon}, \text{ where } \underline{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \text{ and } \underline{\beta} = \begin{bmatrix} a \\ b \end{bmatrix},$$

then $Y_i = a + bx_i + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma^2), i=1,2,\cdots,n$. It follows that $E(Y_i) = a + bx_i$ and $V(Y_i) = \sigma^2$, $i=1,2,\cdots,n$. It can be shown, see Zehna [Ref. 12], that the maximum likelihood estimates of $\underline{\beta}$ and σ^2 are:

$$\hat{\underline{\beta}} = \underline{Q}^{-1} \underline{A}' \underline{Y} \quad \text{and} \quad \hat{\sigma}^2 = \frac{(\underline{Y} - \underline{A}\hat{\underline{\beta}})' (\underline{Y} - \underline{A}\hat{\underline{\beta}})}{n}$$

respectively, where $\underline{Q} = \underline{A}' \underline{A}$. It is further true that $\hat{\hat{\underline{\beta}}}$, the random vector corresponding to $\hat{\underline{\beta}}$ is $N_2(\underline{\beta}, \sigma^2 \underline{Q}^{-1})$ where N_2 indicates a bivariate normal distribution. Algebraically then,

$$\underline{Q} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{bmatrix}, \quad \text{and} \quad \underline{Q}^{-1} = \frac{1}{ns_x^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}.$$

where $s_x^2 = 1/n \sum_{i=1}^{n} (x_i - \bar{x})^2$ and $\bar{x} = 1/n \sum_{i=1}^{n} x_i$.

Then

$$\hat{\underline{\beta}} = \underline{Q}^{-1}\underline{A}'\underline{Y} = 1/ns_x^2 \begin{bmatrix} \bar{y} \sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \end{bmatrix}$$

The above reduces to

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_1^2 - n\bar{x}^2} \tag{2-1}$$

$$\hat{a} = \bar{y} - \bar{x}\hat{b} \tag{2-2}$$

2. <u>Variance of the Estimators</u>

Since $\hat{\hat{\underline{\beta}}}$ is $N_2(\underline{\beta}, \sigma^2 Q^{-1})$ it follows that:

$$E(\hat{\underline{\beta}}) = \underline{\beta}\ ,\ V(\hat{a}) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \bar{x})^2}\ ,\ V(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2-3}$$

and $$COV(\hat{a},\ \hat{b}) = \frac{-\sigma^2\bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

These quantities depend only upon the total number of periods of demand observed and the variance of the underlying distribution in the model. The relative sizes of these variances are indicated in the following table.

TABLE I

VARIANCE OF $\hat{a}$ and $\hat{b}$

n	$V(\hat{a})/\sigma^2$	$V(\hat{b})/\sigma^2$	$COV(\hat{a},\ \hat{b})/\sigma^2$
10	0.4667	0.0121	- 0.0667
50	0.0825	0.0001	- 0.0025
100	0.0406	0.0000	- 0.0006
1000	0.0040	0.0000	- 0.0001

The maximum likelihood estimate of σ^2 is

$$\hat{\sigma}^2 = 1/n \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 .$$

It is well known that

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$$

so that $\frac{n\hat{\hat{\sigma}}^2}{n-2}$ is an unbiased estimator for σ^2 .

3. Prediction

The main purpose of using estimation in inventory applications is to forecast demand at some future period of time. Most often it is necessary to observe demand in periods 1, 2, ···, n, and utilizing these data predict what demand will be in the next period in order to set a reorder level such that within tolerable risk limits the system will not run out of stock.

The demand y_o at time $x_o \neq x_1, \cdots, x_n$, may be considered as a value of the random variable Y_o which is distributed $N(a + bx_o, \sigma^2)$. Then

$$\underset{\sim}{Y}_o = \begin{bmatrix} Y_o \\ \cdots \\ \underset{\sim}{Y} \end{bmatrix} \sim N_{n+1}(\underset{\sim}{A}_o \underset{\sim}{\beta}, \sigma^2 I)$$

where $\underset{\sim}{A}_o = \begin{bmatrix} 1 \ x_o \\ \cdots \\ \underset{\sim}{A} \end{bmatrix}$. It can be shown that

$$Y_o - \hat{\hat{a}} - \hat{\hat{b}}x_o \sim N\left(0, \sigma^2\left[1 + 1/n + \frac{(\bar{x} - x_o)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]\right)$$

and is independent of $n\hat{\hat{\sigma}}^2$. It follows that

$$\frac{\dfrac{Y_o - \hat{\hat{a}} - \hat{\hat{b}}x_o}{\sigma\sqrt{1 + 1/n + \dfrac{(\bar{x} - x_o)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}}}{\sqrt{\dfrac{n\hat{\hat{\sigma}}^2}{\sigma^2(n-2)}}}$$

has a t-distribution with n-2 degrees of freedom. Thus if a reorder level $\hat{R}$ is set by means of

$$\hat{R} = \hat{a} + \hat{b}x_o + t_{n-2,1-\alpha}\, s_{y\cdot x}\sqrt{1 + 1/n + \frac{(\bar{x} - x_o)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \qquad (2\text{-}4)$$

where $\hat{a}$ and $\hat{b}$ are as defined previously,

$$s_{y\cdot x} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}x_2)^2}{n - 2}}\text{, and } t_{n-2,1-\alpha} \text{ is the}$$

$(1-\alpha)$th percentile of the t-distribution with n-2 degrees of freedom, the probability that the demand in period x_o will be less than or equal to $\hat{R}$ is $1-\alpha$. This assumes that a risk of $(100\ \alpha)$% is acceptable.

It should be noted that the length of this one-sided upper prediction limit for Y_o increases as the period for which the forecast is made differs in time from $\bar{x}$. This fact should discourage "Extrapolation" too far above the midpoint of the sampling interval. However, the forecast should be good for the (n+1)st period.

B. CONSTANT MEAN MODEL

The constant mean model may be considered as a special case of the linear model above with b = 0. For this model,

$$Y_i = a + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, \sigma^2) \quad i=1,2,\cdots,n$$

1. Estimates of the Coefficient and Variance

Utilizing the same procedure as before it is found that

$$\hat{a} = 1/n \sum_{i=1}^{n} y_i = \bar{y} \quad , \tag{2-5}$$

$$E(\hat{a}) = a \ , \ V(\hat{a}) = \sigma^2/n \tag{2-6}$$

and $$\hat{\sigma}^2 = 1/n \sum_{i=1}^{n} (y_i - \bar{y})^2$$

2. Prediction

Again it can be shown that $Y_o - \hat{a}$ has a $N(0, \sigma^2[1 + 1/n])$ distribution and is independent of $n\hat{\sigma}^2$. $\frac{n\hat{\sigma}^2}{\sigma^2}$ has a χ^2_{n-2} distribution so that

$$\frac{\dfrac{Y_o - \hat{a}}{\sigma\sqrt{1 + 1/n}}}{\sqrt{\dfrac{n\hat{\sigma}^2}{\sigma^2(n-1)}}}$$ is distributed as t_{n-1} .

It follows that setting a reorder level by means of

$$\hat{R} = \hat{a} + t_{n-1,\ 1-\alpha}\, S_y\sqrt{1 + 1/n} \tag{2-7}$$

where $$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{a})^2}{n - 1}}$$

will yield the desired risk α .

C. LINEAR MODEL WITH Y-INTERCEPT = 0

In many cases it might be reasonable to assume that there is no minimum initial demand imposed on the system, i.e., at time $x = 0$, $Y = 0$, and that demand grows linearly with time. This may be expressed as

$$Y_i = bx_i + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, \sigma^2) \quad i=1,2,\cdots,n.$$

In matrix notation

$$\underline{Y} = \underline{A}b + \underline{\varepsilon} \quad \text{where } \underline{A} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

1. Estimate of Coefficient and Variance

Proceeding as before,

$$\underline{Q} = \underline{A}'\underline{A} = \sum_{i=1}^{n} x_i^2 \qquad \underline{Q}^{-1} = 1/\sum_{i=1}^{n} x_i^2$$

$$\hat{b} = \underline{Q}^{-1}\underline{A}'\underline{Y} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \tag{2-8}$$

$$\hat{\sigma}^2 = (\underline{Y} - \underline{A}\hat{b})' (\underline{Y} - \underline{A}\hat{b}) = 1/n \sum_{i=1}^{n} (y_i - \hat{b}x_i)^2$$

$$V(\hat{\hat{b}}) = \sigma^2 Q^{-1} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2} \qquad E(\hat{\hat{b}}) = b \tag{2-9}$$

2. Prediction

As in the previous case it can be shown that

$Y_o - \hat{\hat{b}}x_o$ has a $N(0, \sigma^2[1 + \frac{x_o^2}{\sum_{i=1}^{n} x_i^2}])$ distribution,

$\frac{n\,\hat{\hat{\sigma}}^2}{\sigma^2}$ is χ^2_{n-1} ,

and further that $Y_o - \hat{\hat{b}}x_o$ and $n\hat{\hat{\sigma}}^2$ are independent.

Thus

$$\frac{\dfrac{Y_o - \hat{b}x_o}{\sigma\sqrt{1 + \dfrac{x_o^2}{\sum_{i=1}^{n}}}}}{\sqrt{\dfrac{n\,\hat{\sigma}^2}{\sigma^2(n-1)}}} \sim t_{n-1}$$

The appropriate reorder level to obtain a risk α is

$$\hat{R} = \hat{b}x_o + t_{n-1,\,1-\alpha} S_{y\cdot x}\sqrt{1 + \frac{x_o^2}{\sum_{i=1}^{n} x_i^2}} \tag{2-10}$$

$$\text{where} \quad S_{y\cdot x} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{b}x_i)^2}{n - 1}}$$

and $t_{n-1,\,1-\alpha}$ is the $(1-\alpha)$th precentile of the t-distribution with n-1 degrees of freedom.

D. CONSTANT MEAN-SIMPLIFIED PROCEDURE

A more easily applied technique for setting the reorder level in the constant mean case was demonstrated by Zehna [Ref. 11] and Ornek [Ref. 6] with very good results. The procedure was used only as an attempt to compare the accuracy of maximum likelihood estimation with that of estimating parameters using exponential smoothing.

Since demand is normally distributed with mean μ and variance σ^2 there is, of course, a theoretical reorder level $\mu + K\sigma$ such that

$$\Pr\,(Y \leq \mu + K\sigma) = 1 - \alpha \tag{2-11}$$

where K is the $(1 - \alpha)$th percentile of the standard normal random variable $- N(0, 1)$.

Hence maximum likelihood methods may be used to estimate μ and σ separately and then these estimates combined as in Eq. (2-11) to set the reorder level. In this way,

$$\hat{\mu} = 1/n \sum_{i=1}^{n} y_i = \bar{y} \qquad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1}}$$

and

$$\hat{R} = \hat{\mu} + K\hat{\sigma} . \tag{2-12}$$

This method of course fails to consider the actual distributions of $\hat{\mu}$ and $\hat{\sigma}$ so that the true risk so obtained is probably not exactly α, but for many purposes the ease of application may make the tradeoff worthwhile.

It is possible to make statements about the expected value and variance of this estimate of reorder level as the distributions of $\hat{\mu}$ and $\hat{\sigma}$ are well known, and the corresponding random variables are further known to be independent.

From Eq. (2-6), $E(\hat{\mu}) = \mu$ and $V(\hat{\mu}) = \sigma^2/n$. It is shown in Hald [Ref. 5] that σ has a Chi-distribution and in particular

$$E(\hat{\sigma}) \doteq a_n \sqrt{1 - \frac{1}{n}} \cdot \sigma \text{ where } a_n = \sqrt{\frac{2}{n - 1}} \quad \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n - 1}{2})}$$

$$\doteq \sqrt{1 - \frac{1}{2(n - 1)}}$$

$$E(\hat{\sigma}) \doteq \sqrt{1 - \frac{3}{2n}}\ \sigma \xrightarrow[n \to \infty]{} \sigma$$

$$V(\hat{\hat{\sigma}}) = \frac{n-1}{n} b_n^2 \sigma^2 \quad \text{where } b_n{}^2 = 1 - a_n^2 \doteq \frac{1}{2(n-1)}$$

$$V(\hat{\hat{\sigma}}) \doteq \frac{1}{2n} \cdot \sigma^2 \xrightarrow[n\to\infty]{} 0 \ .$$

Thus $E(\hat{\hat{R}}) = \mu + K\sqrt{1 - \frac{3}{2n}} \xrightarrow[n\to\infty]{} \mu + K\sigma$,

and $V(\hat{\hat{R}}) = \frac{K^2 + 2}{2n} \sigma^2 \xrightarrow[n\to\infty]{} 0$.

III. EXPONENTIAL SMOOTHING

As indicated in the introduction, exponential smoothing is a method of forecasting discrete time series, in this case future demand, based upon past observations of the system. The major source of information on the development and application of exponential smoothing is the book by Robert G. Brown [Ref. 2]. In fact, Brown introduced the concept and the very name exponential smoothing in an earlier publication. Unless otherwise indicated all of the following discussion of the methodology and theory may be found in the reference noted above.

A. EXPONENTIAL SMOOTHING OF PAST OBSERVATIONS

The basic idea behind exponential smoothing is to assign weights to past observations so as to place the most weight upon recent observations and less and less weight on the older pieces of demand data, thereby limiting the effect on the estimate of "dated" information. This may be accomplished by selecting a smoothing constant $0 < \alpha < 1$ which is exactly the weight placed on the current observation. The smoothing operator is given by

$$\begin{aligned} S_t(Y) &= \alpha \sum_{i=0}^{t-1} \beta^i y_{t-i} + \beta^t y_o \\ &= \alpha y_t + \alpha\beta y_{t-1} + \cdots + \beta^t y_o \end{aligned} \tag{3-1}$$

where $\beta = 1 - \alpha$ and $y_1, y_2, \cdots, y_t$ are the observations of demand in the first t periods, and y_o is some initial demand often taken to be zero. For application in this thesis it will be assumed that $y_o = 0$.

$\tilde{Y} = S_t(Y)$ then is the exponentially smoothed estimate of demand in the (t + 1)st period based upon t periods of observation.

Brown claims that exponential smoothing may be utilized effectively to accurately estimate the coefficients in various models of discrete time series such as demand or sales forecasts. The idea certainly has intuitive appeal and seems to be widely used as a forecasting technique.

For applications of exponential smoothing the number of coefficients to be estimated in a model is referred to as the degrees of freedom of the model. Corresponding to each degree of freedom is an order of exponential smoothing defined by applying the smoothing operator recursively. For single smoothing, as first order smoothing is called, Eq. (3-1) may be written,

$$S_t(Y) = \alpha y_t + \beta S_{t-1}(Y) \tag{3-2}$$

where $S_{t-1}(Y)$ is the current smoothed estimate prior to observing y_t. Nth order smoothing may then be defined as

$$S_t^n(Y) = S[S_t^{n-1}(Y)] = \alpha S_t^{n-1}(Y) + \beta S_{t-1}^n(Y)$$

where $S_t^o(Y)$ is defined to be $S_t(Y)$. As a special case, double smoothing becomes

$$S_t^2(Y) = \alpha S_t(Y) + \beta S_{t-1}^2(Y) \tag{3-3}$$

B. CONSTANT MEAN MODEL -- SINGLE SMOOTHING

Assuming that demand Y is distributed normally with mean μ and variance σ^2 then

$$Y_i = \mu + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, \sigma^2) \quad i=1,2,\cdots,t .$$

Since there is but one coefficient to be estimated, single smoothing is called for to estimate μ. After t periods have been observed the exponentially smoothed estimate of the mean of the demand distribution is given by

$$\tilde{\mu} = S_t(Y) = \alpha y_t + \beta S_{t-1}(Y) = \alpha \sum_{i=0}^{t-1} \beta^i y_{t-i} . \tag{3-3}$$

Note that $E(\tilde{\tilde{\mu}}) = \alpha E(Y) \sum_{i=0}^{t-1} \beta^i = \mu(1 - \beta^t)$, where $\tilde{\tilde{\mu}}$ is the random variable associated with $\tilde{\mu}$, so that, $\lim_{t\to\infty} E(\tilde{\tilde{\mu}}) = \mu$. Thus, $\tilde{\tilde{\mu}}$ is asymptotically unbiased. Also,

$$V(\tilde{\tilde{\mu}}) = \alpha^2\sigma^2 \sum_{i=0}^{t-1} \beta^{2i} = \alpha^2\sigma^2 \left[\frac{1 - \beta^{2t}}{1 + \beta}\right] \tag{3-4}$$

and $\lim_{t\to\infty} V(\tilde{\tilde{\mu}}) = \sigma^2 \left(\frac{\alpha}{1 + \beta}\right)$, a non-zero quantity.

In attempting to estimate the underlying variance of the demand distribution (not considered a coefficient in the model), Brown chooses to use mean absolute deviation, MAD for short, a measure of variability which has questionable theoretical basis and properties [Ref. 10]. MAD is defined as

$$\Delta = E(|Y - \mu|) . \tag{3-5}$$

It is well known that when Y is normally distributed the ratio of Δ to σ is $\sqrt{\frac{2}{\pi}}$, about 0.798. This relationship does not hold however in several other important probability

distributions. To estimate σ Brown first estimates Δ by smoothing forecast errors defined as

$$e(t) = y_t - S_{t-1}(Y) \ , \tag{3-6}$$

the forecast error in period t. The smoothed estimate for Δ is given by

$$\tilde{\Delta}_t = \alpha \sum_{i=0}^{t-1} \beta^i |e_{t-i}|$$

or in operator form

$$\tilde{\Delta}_t = \alpha|e_t| + \beta \tilde{\Delta}_{t-1} \ . \tag{3-7}$$

Then the smoothed estimate of σ is given by

$$\tilde{\sigma} = \sqrt{\frac{\pi}{2}} \sqrt{\frac{2 - \alpha}{2}} \tilde{\Delta} \tag{3-8}$$

implicitly assuming the smoothed estimate for $\tilde{\Delta}$ enjoys the same invariance principle known to be true for maximum likelihood estimates.

It has been clearly pointed out by Zehna [Ref. 11] and others that such a procedure as that described above is statistically suspect. Several attempts have been made to develop the distribution theory necessary to make theoretical judgments concerning $\tilde{\Delta}$ and the invariance principle implicit in obtaining $\tilde{\sigma}$ but so far the mathematics has proven intractible [Ref. 6].

Disregarding the statistical questions involved, Brown combines the smoothed estimates of μ and σ as in Eq. (2-12) to obtain a smoothed estimate of reorder level. In this way,

$$\tilde{R} = \tilde{\mu} + K\tilde{\sigma} \tag{3-9}$$

where K is the appropriate percentile of the N(0, 1) distribution.

The initial condition $S_o(Y)$ is either zero or some best guess as to what the value of μ will be prior to making any observations. The effect of this initial value is negligible after 30-45 periods for $.1 \leq \alpha \leq .9$.

Some remarks may be made concerning the expected value and variance of $\tilde{R}$, the random variable corresponding to $\tilde{R}$,

$$E(\tilde{\tilde{\sigma}}) = \sqrt{\frac{\pi}{2}}\sqrt{\frac{2-\alpha}{2}}\ E(\tilde{\tilde{\Delta}})$$

$$E(\tilde{\tilde{\Delta}}) = \alpha \sum_{i=0}^{t-1} \beta^i E(|\ e_{t-1}\ |).$$

$E(|\ e_{t-1}\ |)$ is shown in [Ref. 11] to be $\sigma\sqrt{\frac{2}{2-\alpha}}\sqrt{\frac{2}{\pi}}$ so that

$$E(\tilde{\tilde{\sigma}}) = (1 - \beta^t)\sigma \xrightarrow[t\to\infty]{} \sigma , \qquad (3\text{-}10)$$

indicating that $\tilde{\tilde{\sigma}}$ is asymptotically unbiased if all the underlying assumptions are valid.

As stated earlier, no analytical results have been obtained concerning the distribution of $\tilde{\Delta}$ and the problems involved extend into the intractibility of obtaining the exact variance of $\tilde{\tilde{\sigma}}$. However, Ornek [Ref. 6] was able to develop a method for approximating $V(\tilde{\tilde{\sigma}})$ numerically. $V(\tilde{\tilde{\sigma}})$ is dependent upon both σ^2 and the value of the smoothing constant α. That is,

$$V(\tilde{\tilde{\sigma}}) = C^2\sigma^2 \qquad (3\text{-}11)$$

where C may be found in the following table.

TABLE II

$\tilde{\tilde{\sigma}}$ VARIANCE FACTORS

α	C	t
0.10	0.1745	t > 80
0.15	0.2173	t > 50
0.20	0.2553	t > 35
0.25	0.2905	t > 30
0.30	0.3239	t > 25
0.35	0.3561	t > 20
0.40	0.3876	t > 15

The values of C were calculated by Ornek [Ref. 6] and the number of periods necessary for a good approximation were determined in the course of computer simulation and testing of the model.

Using the results of Eqs. (3-3) - (3-11) the expected value of $\tilde{\tilde{R}}$ may be written

$$E(\tilde{\tilde{R}}) = (1 - \beta^t)[\mu + K\sigma] \xrightarrow[t\to\infty]{} \mu + K\sigma ,$$

and the variance of $\tilde{\tilde{R}}$ may be approximated as

$$V(\tilde{\tilde{R}}) \doteq [(\frac{\alpha}{1 + \beta})(1 - \beta^{2t}) + C^2K^2]\sigma^2 \tag{3-12}$$

assuming that $\tilde{\tilde{\mu}}$ and $\tilde{\tilde{\sigma}}$ are independent. Again, the variance of $\tilde{\tilde{R}}$ is non-zero even for large t.

C. LINEAR MODEL — DOUBLE SMOOTHING

If the demand in period t is of the form

$$Y_t = a + bt + \varepsilon_t \text{ where } \varepsilon_t \sim N(0, \sigma^2)$$

then $E(Y_i) = a + bi$ and $V(Y_i) = \sigma^2$, $i=1,2,\cdots,t$

1. Estimate of the Coefficients

At time t, the coefficients are estimated by a combination of both single and double smoothing and the demand at time $t + \tau$ is estimated as

$$\tilde{Y}_{t+\tau} = \tilde{a}_t + \tilde{b}_t\tau$$

where τ is the number of periods in the lead time. The estimates of the coefficients developed by Brown are

$$\tilde{a}_t = 2S_t(Y) - S_t^2(Y)$$

$$\tilde{b}_t = \alpha/\beta [S_t(Y) - S_t^2(Y)] \tag{3-13}$$

where $S_t(Y)$ and $S_t^2(Y)$ are given by Eqs. (3-2) and (3-3). It is important to note that $\tilde{a}_t$ is not an estimate of the parameter a in the model but rather an estimate of the current point on the theoretical line at time t. Thus $\tilde{Y}_{t+\tau} = \tilde{a}_t + \tilde{b}_t\tau$ is equal to the estimate of the current theoretical demand increased by the estimated slope multiplied by the number of periods in the lead time. Usually τ will be equal to one and the forecast will be for the next period only.

2. Initial Conditions

The initial conditions for the linear model are

TABLE III

VARIANCE OF SMOOTHING COEFFICIENTS

α	A	B	C
0.10	.12611	.00029	.00539
0.15	.18982	.00107	.01262
0.20	.25377	.00274	.02332
0.25	.31778	.00583	.03790
0.30	.38164	.01099	.05679
0.35	.44508	.01909	.08045
0.40	.50781	.03125	.10938

absolute differences between the estimated point on the line, $\tilde{\mu}_t$, and the actual demand observed in period t.

D. LINEAR MODEL - Y INTERCEPT = 0

When the Y - intercept is known to be zero the model is given by

$$Y_i = bx_i + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, \sigma^2), \quad i=1,2,\cdots,t .$$

Brown does not consider the model as a special case so no smoothed estimate for b is given in [Ref. 2]. Following the general smoothing methodology, it is assumed that only single smoothing is necessary to estimate the single coefficient in the model. At successive periods of time $x_1, x_2, \cdots, x_t$ the corresponding demand $y_1, y_2, \cdots, y_t$ is observed. A natural point estimate of the slope at time x_k would be

$$B_k = y_k/x_k$$

$$S_o(Y) = \tilde{a}_o - (\beta/\alpha)\tilde{b}_o$$

$$S_o^2(Y) = \tilde{a}_o - 2(\beta/\alpha)\tilde{b}_o \qquad (3\text{-}14)$$

where $\tilde{a}_o$ and $\tilde{b}_o$ are initial guesses as to the values of the coefficients a and b. Again these initial conditions are transient and have no effect on the estimate after 30-45 periods.

3. Variance of the Estimators

Using multivariate analysis Brown claims the following asymptotic formulas for the variances of $\tilde{\tilde{a}}$ and $\tilde{\tilde{b}}$ in the linear case.

$$V(\tilde{\tilde{a}}_t) = \frac{\alpha(1 + 4\beta + 5\beta^2)}{(1 + \beta)^3}\,\sigma^2 = A\sigma^2$$

$$V(\tilde{\tilde{b}}_t) = \frac{2\alpha^3}{(1 + \beta)^3}\,\sigma^2 = B\sigma^2$$

$$COV(\tilde{\tilde{a}}_t, \tilde{\tilde{b}}_t) = \frac{\alpha^2(1 + 3\beta)}{(1 + \beta)^3}\,\sigma^2 = C\sigma^2$$

where representative values of A, B, and C are presented in Table III.

4. Setting the Reorder Level

As in the previous case the smoothing estimates of the mean, $\tilde{\mu}_t = \tilde{a}_t + \tilde{b}_t\tau$ and $\tilde{\sigma}$ as defined in Eq. (3-8) may be combined in the form

$$\tilde{R}_t = \tilde{\mu}_t + K\tilde{\sigma}_t \qquad (3\text{-}15)$$

where K is chosen to obtain the desired risk. In this model $\tilde{\Delta}$ is obtained from Eq. (3-7) by smoothing the successive

An estimate of b can then be obtained by smoothing the successive observations of B_k obtaining

$$\tilde{b}_t = S_t(B) = \alpha(y_t/x_t) + \beta S_{t-1}(B) .$$

Since $E(Y_i) = bx_i$, $E(Y_i/x_i) = b$. It follows that

$$E[S_t(B)] = \alpha \sum_{i=0}^{t-1} \beta^i E(Y_{t-i}/x_{t-i}) = (1 - \beta^t)b$$

and $\lim_{t\to\infty} E(\tilde{b}_t) = b$ which implies that this smoothed estimator for b is asymptotically unbiased.

The estimate of σ is obtained as in the previous two models using MAD. In Eqs. (3-7) and (3-8) the forecast error for this procedure may be defined as

$$e_t = | y_t - \tilde{b}_t x_t |$$

IV. COMPARISON OF MLE AND SMOOTHING

In this section the procedures developed in the previous two sections are compared chiefly by means of computer simulation. Before detailing the analysis some general comments on the technique of simulation are in order.

A. COMPUTER SIMULATION

1. General Comments

Computer simulation is primarily utilized to study systems or procedures which cannot be adequately analyzed by means of mathematical analysis or hand calculations. Some advantages of simulation are the rapidity of calculations, magnitude of data which can be generated and analyzed, and the ease with which sensitivity analysis may be conducted. One of the main disadvantages is that the results obtained by simulation are known to be typical of that particular case only and in many cases may not be accurately extended to other sets of circumstances with any assurance that the results are true in general. Also, the results of a simulation study may not be at all applicable to an actual situation if the model developed for the study does not accurately represent the system being analyzed.

2. Use of Random Numbers

Random numbers generated by a computer routine are actually pseudo-random numbers in that the same sequence may be reproduced at will by means of proper initialization of

the sequence. In the strict sense of the word, such replications are not truly random and, therefore, detract from modeling realistic situations such as demand which is known to occur at random. However, it is this property of being reproducible that makes pseudo-random numbers useful in a study which has as its main purpose the comparison of various techniques such as methods of demand forecasting, or the effects of variations in the parameters of the model being studied. To decide if one demand forecasting technique is superior in accuracy to another, the two techniques should be compared under similar circumstances, using the identical demand data. This requirement also is important when the effects on forecasting accuracy of changes in parameters, say the mean or variance of normally distributed demand, are to be measured or tested in some way. If differences in results are to be interpreted as indicating an actual difference in the methods or parameter effects, then the experimenter must insure that such simulation variations are not due simply to the randomness of the data generated.

3. Simulation Used in this Study

As a basis for comparison of maximum likelihood prediction and exponential smoothing as described in Sections II and III, random demand data were generated, computations made according to the method used, and the results compared. The various computer routines utilized were written in Fortran IV and the simulation was conducted using an IBM 360/67 computer. The method of generating the actual data for each specific model is discussed briefly along with the results in

the following sections. It cannot be emphasized enough that all of the numerical results obtained were simulation results and should be interpreted as such. There is no claim made that the results of this thesis are true in general but only that they may give some idea as to properties of the methods tested. The results presented are but a small fraction of the total cases examined and situations studied. Some of the results of the study have been summarized or eliminated entirely for the sake of brevity, but all of the examples presented are fairly typical of the general results obtained.

B. PRELIMINARY CONSIDERATIONS

Before presenting the simulation results a few comments must be made concerning what criteria should be used in judging whether or not a forecasting technique is satisfactory for a particular purpose. Brown [Ref. 2] suggests the following three criteria as a starting point: accuracy; simplicity of computation; flexibility to adjust the rate of response. Accuracy of the forecasting method is discussed in the following section. For the models examined in this section, flexibility of the forecasting technique is not important in that it is assumed that the model is good and that the underlying parameters remain constant over time. Simplicity of computation is discussed in the next paragraph.

One of the main advantages claimed by Brown for exponential smoothing is the ease and simplicity of calculation, to include consideration of the actual computation process and the computer storage required to store the information

necessary to make running computations over time. By formulas (3-2), (3-3), and (3-7) it is easily seen that smoothing requires the retention of only the current value of $S(Y)$ and $\tilde{\Delta}$ for the constant mean case and, in addition, the current value of $S^2(Y)$ for the linear model. At each observation only multiplication and addition are required to update the estimate. Smoothing advocates claim [Ref. 1] that complete historical records must be maintained in order to estimate the parameters in the model by means of maximum likelihood or least squares regression in the linear case. However, it is well known from decision theory that to estimate the parameters in the constant mean case when demand is $N(\mu,\sigma^2)$, the statistics $\sum_{i=1}^{n} Y_i$ and $\sum_{i=1}^{n} Y_i^2$ are sufficient [Ref. 4]. This means that these two statistics are all that must be stored from one period to the next in order to update the estimates with each new observation. Granted, these statistics must be combined at each step with an additional operation or two required including the taking of a square root in the estimation of σ, but with the use of modern high speed computers in most large scale inventory control operations the extra time required for these calculations is probably negligible. For the linear model the statistic $\sum_{i=1}^{n} X_i Y_i$ where X_i is the number of the period in which demand Y_i is observed must also be maintained. The identities which express the formulas in Section II in terms of the above sufficient statistics are developed in Appendix A.

C. COMPARISON BY SIMULATION

For simulation purposes random demand was generated in accordance with the model being studied. Uniform random deviates, U(0,1), were first generated and, recalling the asymptotic normality of the sum of uniform random variables, normal demand was in turn generated. The random demand was tested by several methods to assure that it was indeed normal with the desired parameters. The Chi-Square test (d.f. = 9) was used and the resulting test statistic value of 12.8 fell below the critical value of 21.7 at the 1% level of significance. The normal demand was either used directly in the constant mean case or converted for the linear model by means of an appropriate linear transformation.

Most of the results of the simulation are presented in the form of tables depicting the methods of estimation and the resulting sample quantities of interest. Various streams of random demand were used in many instances from table to table, but within each table the same data were used for each case in order to make suitable comparisons.

1. Simplified MLE and Smoothing in the Constant Mean Model.

Since Zehna [Ref. 11] and Ornek [Ref. 6] previously reported on several aspects of the case where demand is normally distributed with constant mean and demonstrated what appeared to be a significant difference in the accuracy of estimates of the mean obtained by maximum likelihood and exponential smoothing, it was decided to take that model as a point of departure for the simulation study in this thesis.

In this section various results of forecasting demand using the simplified maximum likelihood approach Eq. (2-12) and smoothing Eq. (3-9) are presented.

As mentioned in the introduction, the approach taken was to generate random demand, make estimates of the various parameters after particular sampling intervals, replicate the experiment a number of times, and then determine the average values of the estimates in the particular periods and the variability of the estimators over the replicated experiments. Table IV presents these results for estimates of reorder level after 15, 30, 45, and 300 periods of demand observation over 100 replications. The parameter pairs (μ, σ) are of the same proportions as those studied by Zehna and previously indicated to be realistic by NavSup. The theoretical reorder level was determined from Eq. (2-11) with the risk level constant at 0.05 for each parameter pair. In this case, and throughout the study unless otherwise indicated, the initial values for smoothing were chosen in order to give smoothing the best possible starting conditions, that is,

$$S_o(Y) = \mu, \text{ and } \Delta_o(Y) = \sqrt{\frac{2}{\pi}}\sqrt{\frac{2}{2-\alpha}}\,\sigma. \tag{4-1}$$

A smoothing constant value of $\alpha = 0.20$ was used, the value presently being used by NavSup in their forecasting models.

The entries in Table IV are the average estimate of reorder level and the sample standard deviation of the estimators over 100 replications for the indicated parameter pairs and number of periods. The results are rather striking and consistent. In most cases exponential smoothing tends to

TABLE IV

ESTIMATE OF REORDER LEVEL, $\hat{R}$ AND $\tilde{R}$

FOR VARIOUS PARAMETER PAIRS

Risk = 0.05	Theoretical Reorder Level	15 Periods		30 Periods		45 Periods		300 Periods	
Parameters		$\hat{R}$	$\tilde{R}$	$\hat{R}$	$\tilde{R}$	$\hat{R}$	$\tilde{R}$	$\hat{R}$	$\tilde{R}$
(10,1)	11 64	11.44 0.38	12.69 0.70	11.54 0.23	11.75 0.48	11.55 0.20	11.58 0.52	11.63 0.08	11.73 0.54
(20,2)	23.29	22.87 0.76	25.39 1.40	23.07 0.46	23.49 0.95	23.10 0.39	23.16 1.03	23.26 0.17	23.47 1.09
(30,3)	34.93	34.31 1.15	38.08 2.10	34.61 0.70	35.24 1.43	34.66 0.59	34.73 1.55	34.89 0.25	35.20 1.63
(40,4)	46.58	45.75 1.53	50.77 2.80	46.14 0.93	46.98 1.90	46.21 0.79	46.31 2.06	46.52 0.34	46.93 2.17
(50,5)	58.22	57.19 1.91	63.46 3.50	57.68 1.16	58.73 2.38	57.76 0.99	57.89 2.58	58.15 0.42	58.66 2.71
(100,10)	116.45	114.37 3.82	126.93 6.99	115.36 2.32	117.46 4.76	115.75 1.97	115.78 5.15	116.30 0.84	117.33 5.43
(500,50)	582.25	571.87 19.12	634.64 34.96	576.79 11.59	587.31 23.79	577.62 9.86	578.88 25.77	581.50 4.19	586.64 27.14
(1000,100)	1164.50	1143.73 38.24	1269.26 69.92	1153.56 23.18	1174.61 47.57	1155.23 19.71	1157.78 51.55	1163.00 8.39	1173.26 54.28

overestimate the reorder level while MLE underestimates, a result noted also by Zehna. However, the most noteworthy aspect is the much larger variability of the smoothing estimate demonstrated by a relatively higher sample standard deviation. As more and more observations are taken the sample standard deviation of the maximum likelihood estimate decreases, a fact that is easily verified in the theory. However, this property is not demonstrated by the smoothing estimates. This point was suggested by Ornek [Ref. 6] and indicated by Eq. (3-12). In fact, after 300 periods the sample standard deviation of smoothing is consistently over six times as large as that of maximum likelihood.

It is indicated in Table IV that for these particular parameter pairs the results do not depend, on the average, upon the various values of the parameters. For example, the average reorder level estimate given by MLE after 15 periods for the pair (10,1) is exactly 1/10th the average estimate for the pair (100,10). This result occurred in every case examined. This is not surprising since the random demand for each case was generated by the identical stream of random numbers, altered only by the specific parameters. For this reason it was decided to consider further only the parameter pair (100,10) as a typical case. Continuing with this approach, Table V presents the results for the (100,10) parameter pair with estimates being made after 5 to 100 periods. The entries are the average reorder level and sample standard deviation of the estimates over 100 replications. These results are consistent with the ideas brought out by Table IV,

TABLE V

ESTIMATE OF REORDER LEVEL
FOR (100, 10) CASE AND VARIOUS RISKS

Risk & AROLV*	Estimator	Number of Periods							
		5	10	20	30	40	50	75	100
0.01	$\hat{R}$	115.66 8.53	118.74 5.30	121.02 3.77	121.63 3.14	121.98 2.53	122.17 2.38	122.44 2.02	122.58 1.73
123.30	$\tilde{R}$ $\alpha = 0.2$	122.87 7.40	122.45 5.72	123.86 6.55	123.17 7.10	122.90 6.78	123.28 7.52	123.19 6.46	123.39 7.17
0.05	$\hat{R}$	110.90 6.89	113.20 4.36	114.87 2.98	115.32 2.46	115.54 2.02	115.67 1.93	115.88 1.66	115.97 1.45
116.45	$\tilde{R}$ $\alpha = 0.2$	116.05 5.76	115.88 4.68	116.83 5.09	116.43 5.60	116.77 5.47	116.41 5.90	116.42 5.21	116.52 5.69
0.10	$\hat{R}$	108.37 6.12	110.26 3.93	111.61 2.61	111.97 2.14	112.12 1.80	112.22 1.72	112.40 1.49	112.46 1.33
112.82	$\tilde{R}$ $\alpha = 0.2$	112.44 4.97	112.39 4.23	112.10 4.39	112.86 4.87	112.60 4.85	112.76 5.12	112.84 4.64	112.89 4.97
0.25	$\hat{R}$	104.16 5.08	105.35 3.41	106.16 2.15	106.37 1.74	106.41 1.52	106.45 1.48	106.58 1.26	106.60 1.16
106.75	$\tilde{R}$ $\alpha = 0.2$	106.40 3.88	106.56 3.69	106.87 3.47	106.88 3.83	106.63 4.05	106.66 4.08	106.84 3.95	106.80 3.95

* Actual Reorder Level

TABLE V--Continued

Risk & AROLV*	Estimator	Number of Periods							
		5	10	20	30	40	50	75	100
0.50	$\hat{R}$	99.46 4.50	99.89 3.23	100.10 1.99	100.15 1.59	100.06 1.44	100.04 1.39	100.11 1.14	100.09 1.08
100.00	$\tilde{R}$ $\alpha = 0.2$	99.68 3.31	100.08 3.54	99.94 3.09	100.24 3.20	99.99 3.66	99.88 3.58	100.17 3.73	100.04 3.30

* Actual Reorder Level

but some additional insight into the source of reorder level estimate variability may be gained by examining Table V. After 100 periods the sample standard deviation of $\tilde{R}$ for risk level 0.01 is over twice as large as the same statistic for a risk of 0.50. This is because in the estimate at the former risk level, considerable weight is given to the estimate of σ while at the latter risk level, $K = 0$ and the reorder level estimate is determined solely by $\tilde{\mu}$. Thus, the weight given to $\tilde{\sigma}$ in the estimate decreases as the risk increases from 0.01 to 0.50 and the corresponding variability of $\tilde{R}$ decreases. The same general effect is true for $\hat{R}$ but to a lesser extent. This is to be expected since, from the theoretical development, $\tilde{\sigma}$ is roughly twice as variable as $\tilde{\mu}$ in the limiting case (2.3 times for $\alpha = 0.2$). For MLE the variance of $\hat{\sigma}$ is half that of $\hat{\mu}$.

To determine the effects of staring conditions, or initial values, on the ultimate estimate given by smoothing the same experiment was conducted comparing the perfect initial values given in Eq. (4-1) with $S_o(Y) = 0$ and $\Delta_o(Y) = 0$. The results are listed in Table VI for various values of the smoothing constant α to determine how long the initial values effect the reorder level estimate and the variance of the estimate. In the Table, $\tilde{R}1$ implies the zero initial values and $\tilde{R}2$ indicates the best possible starting conditions. The approximate period in which the effect of the initial values become negligible is indicated as the "catch-up period." One of the most interesting results of this Table is that the

TABLE VI

EFFECT OF INITIAL CONDITIONS
FOR SMOOTHING WITH VARIOUS SMOOTHING CONSTANTS

α	CUP*	Estimator	Number of Periods: 5	10	20	40	75	100
0.05	Over 100	$\tilde{R}1$	63.76	104.27	141.13	142.50	122.68	118.26
			2.95	3.58	3.36	3.95	2.94	2.68
		$\tilde{R}2$	116.73	116.25	116.59	116.47	116.47	116.48
			1.72	1.94	2.18	2.42	2.62	2.57
0.10	100	$\tilde{R}1$	106.37	143.12	142.58	120.85	116.60	116.56
			5.16	5.60	4.82	4.18	3.61	3.88
		$\tilde{R}2$	116.21	116.09	116.73	116.33	116.48	116.55
			3.24	3.21	3.40	3.61	3.60	3.88
0.15	65	$\tilde{R}1$	132.66	149.51	128.14	116.70	116.45	116.55
			6.92	7.00	5.31	4.65	4.46	4.88
		$\tilde{R}2$	116.12	115.98	116.81	116.22	116.45	116.55
			4.58	4.08	4.32	4.58	4.46	4.88
0.175	55	$\tilde{R}1$	140.98	140.98	123.60	116.34	116.43	116.54
			7.70	7.38	5.49	5.06	4.85	5.30
		$\tilde{R}2$	116.08	115.92	116.83	116.18	116.43	116.54
			5.19	4.41	4.72	5.04	4.85	5.30
0.20	45	$\tilde{R}1$	146.69	142.92	120.72	116.21	116.42	116.52
			8.43	7.44	5.59	5.47	5.21	5.69
		$\tilde{R}2$	116.05	115.88	116.83	116.17	116.42	116.52
			5.76	4.68	5.09	5.47	5.21	5.69
0.275	40	$\tilde{R}1$	150.18	138.15	118.96	116.18	116.42	116.50
			9.13	7.50	5.73	5.89	5.54	6.05
		$\tilde{R}2$	116.02	115.83	116.82	116.17	116.42	116.50
			6.70	4.92	5.42	5.89	5.54	6.05
0.25	35	$\tilde{R}1$	151.83	133.48	117.94	116.17	116.40	116.47
			9.81	7.49	5.52	6.30	5.86	6.38
		$\tilde{R}2$	115.99	115.78	116.80	116.17	116.40	116.47
			6.80	5.13	5.73	6.30	5.86	6.38

* "Catch-Up Period"

choice of smoothing constant α has little effect on the average value of the reorder level estimate but the standard deviation of the estimate increases with increasing α. This is not surprising since for a stable system, as in the current case, the smaller values of the smoothing constant are clearly indicated. This choice of α causes the resulting estimator to be less sensitive to changes in the system, but have much smaller variance than for larger values of the smoothing constant.

One possible measure of the accuracy of an estimator, at least for comparison purposes, is the mean squared error of the estimate. Mean squared error (MSE) may be defined as the sum of the square of the bias and the variance. The theoretical MSE and average sample MSE for the (100,10) parameter pair after 200 observations of demand are displayed in Table VII for the MLE and smoothing estimates of reorder level. The sample results compare favorably with the theoretical values which for smoothing are approximations based upon the results of Section III. Not surprisingly, the results favor the MLE procedure, due in large part to the much smaller variance. In fact, the variance in both cases is the dominating factor as the bias is negligible after 200 periods.

As one further illustration of the greater sample variance of $\tilde{R}$, reorder level for the (100,10) case was estimated using 1000 periods of demand observations and the experiment repeated 100 times. A graph of the MLE and smoothing

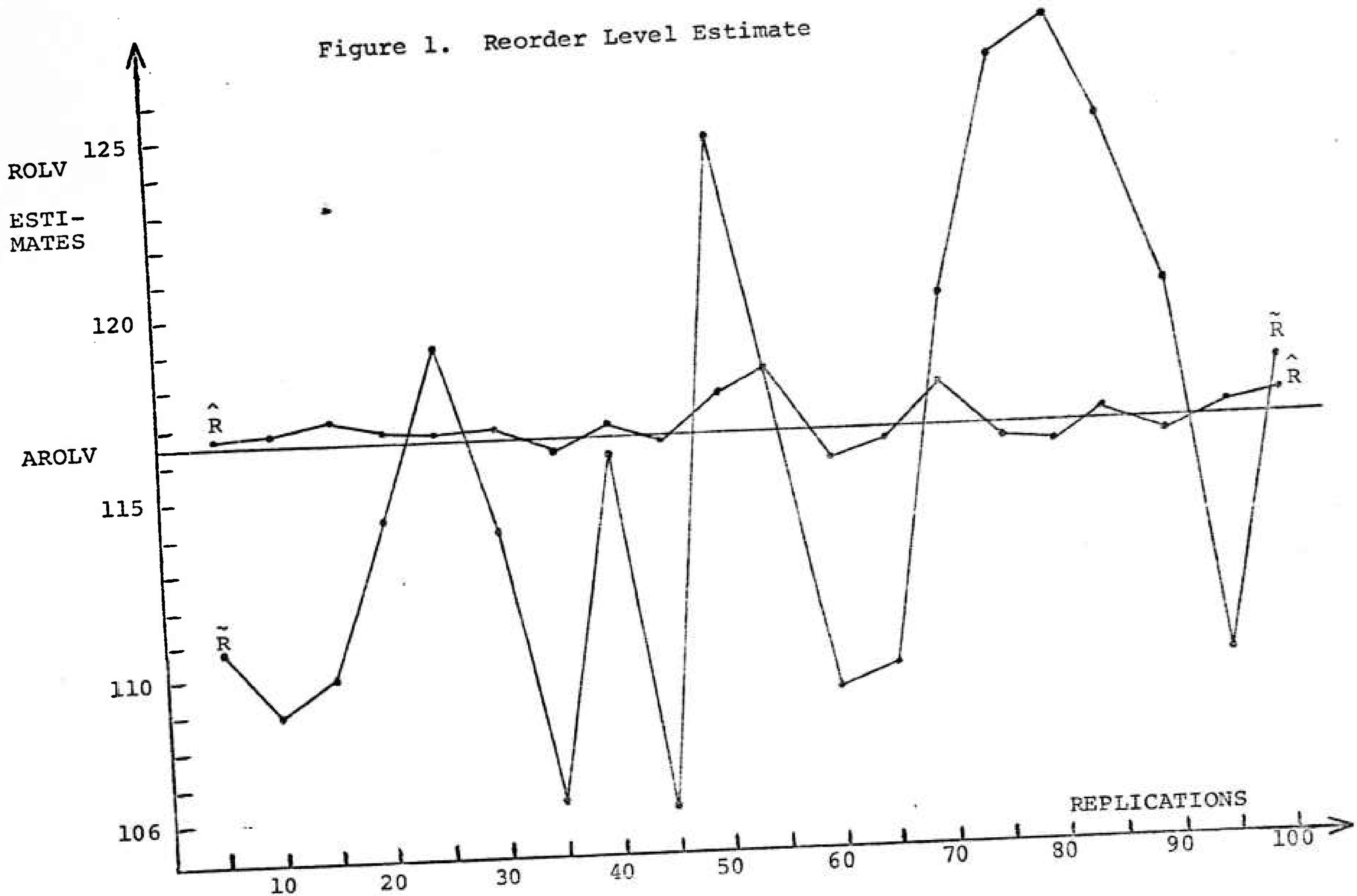
TABLE VII

THEORETICAL AND SAMPLE MEAN SQUARED
ERROR IN PERIOD 200

Theoretical Mean Squared Error								
Risk	MLE	Smoothing with Indicated α						
		0.10	0.15	0.20	0.25	0.30	0.35	0.40
0.01	1.87	21.79	33.74	46.50	60.10	74.60	90.05	106.56
0.05	1.13	13.50	20.89	28.75	37.12	46.04	55.53	65.65
0.10	0.91	10.27	15.87	21.82	28.16	34.89	42.05	49.69
0.25	0.62	6.65	10.26	14.08	18.13	22.43	26.99	31.35
0.50	0.50	5.26	8.11	11.11	14.29	17.65	21.21	25.00
Average Sample Mean Squared Error over 100 Replications								
Risk	MLE	Smoothing with Indicated α						
		0.10	0.15	0.20	0.25	0.30	0.35	0.40
0.01	1.64	22.94	35.64	50.55	67.26	84.72	102.28	119.63
0.05	1.01	14.06	22.29	32.01	42.99	54.56	66.24	77.84
0.10	0.78	10.57	17.01	24.64	33.25	42.37	51.60	60.82
0.25	0.53	6.62	10.98	16.10	21.82	27.87	34.04	40.24
0.50	0.47	4.99	8.37	12.16	16.21	20.40	24.66	28.99

estimates of reorder level over the 100 replications is displayed in Fig. 1. The theoretical reorder level for a risk of 0.05 is 116.45. Although the average values of $\tilde{R}$ and $\hat{R}$ are not significantly different, the difference in variability is quite obvious. In fact, even after the 1000 periods the sample standard deviation of $\tilde{R}$ is 5.654, while that of $\hat{R}$ is but 1.007. In view of many other experiments of this

Figure 1. Reorder Level Estimate



type that were conducted, it appears that for a risk level of 0.05 and $\alpha = 0.2$, a lower bound on the standard deviation of $\tilde{R}$ is approximately $\sigma/2$. No results to that effect have been determined analytically so the remark is included here only as a matter of interest.

As a final example of results under this particular case, a comparison of the sample risks obtained by both methods of estimation was made. Again with $\mu = 100$ and $\sigma = 10$ demand was generated and the reorder level estimated after various sampling intervals. This time the experiment was repeated 500 times and the risks over the 500 replications in the indicated periods are presented in Table VIII. In several of the cases the results do not differ significantly in one way or the other, but generally both methods tend to yield higher risk than desired. In the case of smoothing this is due to the variability of the estimate, and in the case of MLE the higher risk may be attributed to the tendency to underestimate reorder level. The risk obtained by smoothing was greater than that obtained by MLE in the majority of instances.

2. Linear Model

In this section a study similar to that of the previous section is presented for the linear model. The exact results of the maximum likelihood development of Section II are used for comparison with the exponentially smoothed estimates. The constant mean and zero Y-intercept cases are presented as special cases of the general linear model.

TABLE VIII

SAMPLE RISK OBTAINED BY SMOOTHING AND SIMPLIFIED MLE PROCEDURE

Desired Risk	Estimator	Number of Periods 10	20	50	100	150	200
0.01	$\hat{R}$	.044	.016	.018	.014	.006	.010
	$\tilde{R}$	.026	.024	.036	.030	.014	.032
0.05	$\hat{R}$	.078	.076	.062	.064	.046	.060
	$\tilde{R}$	.088	.078	.070	.096	.064	.076
0.10	$\hat{R}$	.134	.132	.120	.134	.086	.108
	$\tilde{R}$	.120	.120	.114	.142	.110	.128
0.25	$\hat{R}$	.272	.270	.248	.246	.224	.250
	$\tilde{R}$	.264	.284	.258	.272	.256	.250
0.50	$\hat{R}$	.514	.480	.504	.494	.500	.482
	$\tilde{R}$	.532	.480	.496	.474	.488	.472

The first simulation experiment performed in this section was to generate random demand according to the linear model given by

$$y = a + bx + \varepsilon , \quad \text{where} \quad \varepsilon \sim N(0,\sigma^2)$$

for various values of a, b, and σ. The maximum likelihood and smoothing estimates of the parameters in the model were calculated after 100 periods of demand observation and the results averaged over 100 replications. The estimate along with the sample standard deviation of the estimates are displayed in Table IX. As indicated in Section III the smoothing method does not estimate the value of the intercept separately but rather the point on the theoretical line at each period. For purposes of the Table only, $\tilde{a}_t$ was

$\tilde{b}_t x_t$ from the average smoothed estimate of the point on the line for each period of interest, and thus no sample standard deviation of a_t was obtained. The results of Table IX are

TABLE IX

ESTIMATE OF PARAMETERS

IN A LINEAR MODEL AFTER 200 PERIODS

Parameters	MLE		Smoothing	
(a,b,σ)	Sample Average	Sample Standard Deviation	Sample Average	Sample Standard Deviation
(5,.1,1)				
a	4.84	.095	5.23	*
b	.10	.025	.10	.050
σ	.93	.077	1.04	.294
(50,.2,3)				
a	50.57	.378	50.64	*
b	.20	.099	.19	.151
σ	2.85	.200	3.13	.882
(100,.3,4)				
a	100.24	.698	100.87	*
b	.30	.117	.30	.201
σ	3.96	.347	4.17	1.176
(500,.5,5)				
a	499.68	.540	501.05	*
b	.50	.249	.50	.252
σ	4.96	.377	5.21	1.470

* See Narrative

not very revealing except that once again the smoothed estimates are more variable than the corresponding maximum likelihood estimates of the parameters. More interesting was

the fact that for many experiments using various values of a, b, and σ the sample risks obtained by both methods were seen to be independent of the particular parameters in the model, and appear to depend only upon the number of observations upon which the estimates are based. For this reason it was decided to again direct attention to only a couple of parameter values.

For the sake of brevity only a few of the experiments conducted are reported here, in that the results are much the same as in the previous sections. Not much insight could be gained from attempting to analyze the sample estimates of the parameters. It appeared more beneficial to examine the sample risks obtained by the two methods of estimation. Table X presents the sample risks obtained after the indicated number of periods over 1000 replications. It should be pointed out that the entries in the Table are not the fraction of time the demand exceeded the reorder level estimates through 50 periods for example, as was the case in the results presented by Zehna and Ornek, but rather the percentage of times the actual demand exceeded the reorder level in the (p+1)st period over the 1000 samples of the estimates. The reasons for proceeding in this manner were discussed in the introduction. For this case the parameters (a,b,σ) are (50,2,5). Again the best possible initial values were used for smoothing, namely

$$S_0(Y) = a - (\alpha/\beta)b \text{ , and } S_0^2(Y) = a - 2(\alpha/\beta)b \text{ .}$$

TABLE X

LINEAR MODEL – ACTUAL RISK OBTAINED

IN 1000 REPLICATIONS ($a=50$, $b=2$, $\sigma=5$)

Desired Risk	Number of Periods							
	10		20		50		100	
	$\hat{R}$	$\tilde{R}$	$\hat{R}$	$\tilde{R}$	$\hat{R}$	$\tilde{R}$	$\hat{R}$	$\tilde{R}$
0.01	.007	.036	.014	.027	.011	.024	.008	.029
0.05	.045	.103	.045	.070	.049	.078	.053	.068
0.10	.092	.155	.098	.124	.103	.137	.110	.114
0.25	.233	.306	.256	.264	.249	.274	.256	.249
0.50	.482	.516	.521	.519	.510	.536	.505	.517

The results here are noteworthy in that the sample risk obtained by the smoothed estimates exceeded that of MLE in almost every case, in the sense that the smoothing risk differed more in one way or the other from the desired risk than did the risk obtained by MLE. It is felt that a sample size of 1000 for each period makes the results fairly significant, but once again it must be recognized that the results are sample values from simulation and should be judged accordingly.

Since the simplified maximum likelihood procedure yielded good results in the constant mean case, it was decided to try a similar version for the linear case. To apply the methods of Section II exactly, and thereby be able to make exact probability statements about risk, the value of

the t-distribution percentile must be changed with each additional period of demand observation. As n, the total number of periods, becomes large, however, the value of the t-distribution percentiles approach those of a standard normal distribution, indicated by K throughout this thesis. In view of this, it was decided to determine what effect replacing the varying values of t with the corresponding limiting value of K, always smaller, would have on the sample risk obtained by MLE. The resulting estimate is referred to as the K-version in this report. In Table XI this has been done for the general linear case with parameter triple (100,2,10). The Table entries are the average reorder level estimate, sample standard deviation of the estimates over 200 replications, and the sample risk obtained by the indicated method of estimation. The desired risk levels of 0.25 and 0.50 have been omitted primarily for the sake of brevity once again, but also because the values are probably too large for any practical application. Tables XII and XIII present the same information for the special cases (100,0,10) and (0,2,10) respectively.

The results of these final three tables for this section are consistent with the earlier results. The sample standard deviation of the smoothed estimates is relatively constant over the number of periods and the corresponding values for both versions of MLE, although consistently larger in the early periods, decrease over time and are roughly a third as large as the deviation of the corresponding smoothed estimates after 100 periods. The maximum likelihood risk

TABLE XI

REORDER LEVEL, SAMPLE STANDARD DEVIATION OF ESTIMATOR, AND SAMPLE RISK OVER 200 REPLICATIONS. LINEAR (100,2,10)

Desired Risk	Estimator	Number of Periods 10	50	100
0.01	$\hat{R}$ - T	156.63,10.62 .02	227.41,3.77 .01	326.11,2.62 .01
	$\hat{R}$ - K	150.67,9.59 .02	226.59,3.72 .01	325.71,2.60 .01
	$\tilde{R}$	145.83,8.54 .03	228.11,8.58 .01	327.60,8.65 .02
0.05	$\hat{R}$ - T	144.64,8.67 .04	219.74,3.34 .05	318.87,2.31 .04
	$\hat{R}$ - K	142.36,8.36 .05	219.42,3.32 .06	318.70,2.30 .05
	$\tilde{R}$	138.92,7.26 .08	220.51,7.24 .07	319.98,7.16 .06
0.10	$\hat{R}$ - T	139.19,7.97 .08	215.81,3.16 .08	315.07,2.18 .13
	$\hat{R}$ - K	137.96,7.84 .09	215.63,3.15 .09	314.99,2.18 .13
	$\tilde{R}$	135.27,6.67 .13	216.48,6.66 .11	315.94,6.49 .12

TABLE XII

REORDER LEVEL, SAMPLE STANDARD DEVIATION OF ESTIMATOR, AND SAMPLE RISK OVER 200 REPLICATIONS. LINEAR (100,0,10)

Desired Risk	Estimator	Number of Periods		
		10	50	100
0.01	$\hat{R}$ - T	129.78,7.29 .01	124.48,2.76 .01	123.93,1.90 .01
	$\hat{R}$ - K	124.61,6.28 .01	123.69,2.69 .01	123.54,1.87 .01
	$\tilde{R}$	121.85,6.71 .01	124.16,6.98 .03	12[illegible].73,6.86 .02
0.05	$\hat{R}$ - T	119.37,5.32 .04	117.04,2.14 .05	116.78,1.48 .04
	$\hat{R}$ - K	117.39,4.97 .05	116.74,2.11 .05	116.64,1.47 .04
	$\tilde{R}$	115.48,5.36 .10	117.15,5.44 .08	116.71,5.36 .07
0.10	$\hat{R}$ - T	114.62,4.51 .09	113.23,1.85 .11	113.06,1.29 .09
	$\hat{R}$ - K	113.56,4.35 .13	113.06,1.84 .11	112.98,1.28 .10
	$\tilde{R}$	112.10,4.71 .13	113.43,4.72 .13	112.99,4.65 .11

TABLE XIII

REORDER LEVEL, SAMPLE STANDARD DEVIATION OF ESTIMATOR, AND SAMPLE RISK OVER 200 REPLICATIONS. LINEAR (0,2,10)

Desired Risk	Estimator	Number of periods		
		10	50	100
0.01	$\hat{R}$ - T	54.62,9.25 .01	127.04,3.42 .01	226.08,2.39 .01
	$\hat{R}$ - K	48.12,8.10 .02	125.96,3.34 .02	225.52,2.36 .01
	$\tilde{R}$	47.62,8.80 .01	126.40,7.11 .03	225.80,6.93 .01
0.05	$\hat{R}$ - T	43.28,7.43 .04	119.46,2.93 .05	218.88,2.06 .04
	$\hat{R}$ - K	40.79,6.98 .06	118.97,2.90 .05	218.61,2.05 .04
	$\tilde{R}$	40.74,7.32 .06	119.32,5.59 .08	218.76,5.44 .07
0.10	$\hat{R}$ - T	38.11,6.74 .08	115.58,2.72 .10	215.11,1.92 .11
	$\hat{R}$ - K	36.90,6.46 .09	115.27,2.70 .10	214.95,1.92 .13
	$\tilde{R}$	37.09,6.61 .10	115.57,4.88 .13	215.03,4.73 .11

using the exact t factors is never larger than the risk when the corresponding values of K are used. On the basis of these results, it would be difficult to state that the modified K-version of maximum likelihood estimation of re-order level is significantly poorer than the actual estimator using t. In fact, in the earlier periods the sample standard deviation of the K-version is smaller than that of the actual MLE estimation technique. This of course is due to the smaller weight placed upon the estimate of σ in the earlier periods.

V. BAYES PROCEDURES

At times the manager of a supply system may have "a priori" degrees of belief about the values of the unknown parameters in the demand distribution assumed for a particular model. In other words, he may subjectively feel, prior to actually observing any demand data, that certain values of the unknown parameters in the model are more likely than others. He may further wish to include his subjective feelings along with observed values of the random demand to make estimates of the unknown parameters. Bayesian methods may accomplish the above, combining the real data and subjective feelings into estimates which actually minimize the decision maker's risk as defined in a very special way.

The subjective probabilities which measure the degree of belief discussed above determine a prior distribution for the parameters. The sample values of the observed demand are used with the prior distribution to determine a posterior distribution which is then used in order to obtain an estimate of the unknown parameters. The choice of a prior distribution for the unknown parameters is certainly arbitrary and has lead to general controversy as to the appropriateness and value of Bayesian methods. It is not much more arbitrary, however, than the selection of initial values for use with exponential smoothing estimation techniques. These initial values were shown in Section IV to effect the estimate for as many as 100 periods, depending upon the value of the smoothing

constant, another aspect of smoothing which is certainly open to subjectivity. Without attempting to discuss the appropriateness of Bayesian methods for the demand forecasting problem at hand, and certainly without recommending its use without further examination, the concept is developed in this section as it applies to the model of normal demand with constant mean. The Bayes estimation results are then compared to exponential smoothing in a variety of situations.

A. BAYES ESTIMATION

One of the many limitations of Bayesian procedures is the extreme difficulty of obtaining joint prior and posterior distributions for estimating more than a single parameter in a model. However unrealistic it may appear, it will be assumed here that demand is normally distributed with mean θ and <u>known</u> variance σ^2. This might even be plausible in cases where the demand for various items of supply demonstrates about the same degree of variability but fluctuates about different and unknown mean values. In any case, it is assumed that

$$f(y|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2}$$

where $f(y|\theta)$ indicates that the distribution of Y is conditioned on the value of the parameter θ. It can be shown [Ref. 8] that if the prior distribution assumed for θ is $N(\theta_o,\sigma_o^2)$, then the posterior distribution is also normal with mean

$$\frac{\sigma_o^2 y + \sigma^2 \theta_o}{\sigma_o^2 + \sigma^2} \tag{5-1}$$

and variance

$$\frac{\sigma^2 \sigma_o^2}{\sigma_o^2 + \sigma^2} \tag{5-2}$$

Thus the mean of the posterior distribution is a linear combination of the observed demand and the mean of the prior distribution θ_o. For estimation purposes it can be shown that the mean of the posterior distribution as an estimate of the parameter θ minimizes the loss as measured by squared error. That is, if θ' is an estimate of θ and the conditional and prior distributions are as specified above, then $(\theta - \theta')^2$ is minimized by choosing θ' equal to the mean of the posterior distribution.

It should be noted that in Eq. (5-1) the weight given to the observation in the linear combination is $\frac{\sigma_o^2}{\sigma^2 + \sigma_o^2}$ so that if the decision maker is fairly certain that θ_o is the true value of the parameter θ or if he wishes to bias the estimate in favor of his prior feeling, he should choose a small variance for the prior. If the exact value of θ is less certain however, the variance of the prior should be relatively large compared to θ^2. In this way, most of the weight in the resulting estimate will be placed on the observed demand.

B. BAYES FORECAST MODEL

The Bayes estimate of mean demand developed in the preceding section may be used to forecast demand and set an appropriate reorder level to obtain an approximate risk in a manner analogous to the procedures used in Section II-D. The demand in period one is observed and assuming a $N(\theta_o, \sigma_o^2)$ prior distribution on θ, a posterior distribution of θ given y_1 may be determined. As previously derived, the posterior is normal with mean

$$\theta^* = \frac{\sigma_o^2 y_1 + \sigma^2 \theta_o}{\sigma_o^2 + \sigma^2}$$

and variance

$$\sigma_1^2 = \frac{\sigma_o^2 \sigma^2}{\sigma_o^2 + \sigma^2} .$$

The estimate of reorder level then is given by

$$R_1^* = \theta_1^* + K\sigma .$$

Since information is gained by the observation y_1, the posterior distribution determined in period one is used as the prior distribution for period two. After observing y_2, a new posterior distribution for θ is determined. It is again normal with mean

$$\theta_2^* = \frac{\sigma_1^2 y_2 + \sigma^2 \theta_2^*}{\sigma_1^2 + \sigma^2}$$

and variance

$$\frac{\sigma_1^2 \sigma^2}{\sigma_1^2 + \sigma^2} .$$

Upon substitution for θ_1^* and σ_1^2 the above becomes

$$\theta_2^* = \frac{\sigma_c^2(y_1+y_2) + \sigma^2\theta_o}{2\sigma_o^2 + \sigma^2} \quad \text{and} \quad \sigma_2^2 = \frac{\sigma^2\sigma_o^2}{2\sigma_o^2 + \sigma^2} .$$

Continuing in this manner it may be shown by induction that after n periods the estimate and variance are

$$\theta_n^* = \sigma_o^2 \sum_{i=1}^{n} y_i + \sigma^2\theta_o$$

and

$$\sigma_n^2 = \frac{\sigma_o^2\sigma^2}{n\,\sigma_o^2 + \sigma^2} .$$

It follows that

$$E(\Theta_n^*) = \frac{n\sigma_o^2\,\theta + \sigma^2\theta_o}{n\,\sigma_o^2 + \sigma^2}$$

$$\lim_{n\to\infty} E(\Theta_n^*) = \theta$$

and

$$\lim_{n\to\infty} \sigma_n^2 = 0 .$$

The reorder level set for the (n+1)st period is given by

$$R_n^* = \theta_n^* + K\sigma$$

where K is the $(1-\alpha)$th percentile of the N(0,1) distribution.

C. SELECTION OF PARAMETERS FOR THE PRIOR--SIMULATION RESULTS

The results of Bayes estimation depend strongly in most cases upon the choice of the parameters in the prior distribution. The ratio of σ_o^2 to σ^2 determines the weight given to the observed demand relative to the weight placed upon the "a priori" best guess of mean demand θ_o. To determine exactly how much the estimate of mean demand is effected by various combinations of θ_o and σ_o^2, a limited amount of simulation was

conducted. For the constant mean case with θ = 100 and σ = 10, random demand was generated for 100 periods. The Bayes estimate of θ was computed for combinations of θ_o = 0, θ, $\theta/2$, $\theta/3$, $\theta/4$, $\theta/5$, and $\sigma_o^2 = \sigma$, $\sigma^2/2$, $\sigma^2/3$, $\sigma^2/4$, $\sigma^2/5$, $2\sigma^2$, $3\sigma^2$, $4\sigma^2$, $5\sigma^2$. The results of this investigation are summarized in Table XIV, the entries being the Bayes estimates of θ after 5 and 100 periods of demand observation. Obviously the closer θ_o is to the true value of θ the better the resulting estimate is. It should be noted however, that for this ratio of σ_o^2/σ^2 sufficiently large, the Bayes estimate approaches the actual value of θ rapidly. For example, in the above case with θ_o = 0 and $\sigma_o^2 = 5\sigma^2$ the Bayes estimate of θ was 96.6 after just 5 periods and 99.5 at the end of 50 periods.

From the detailed data obtained it appeared that the Bayes estimate was fairly stable especially for a relatively good selection of θ_o. For this reason it was decided to compare the Bayes estimates with the corresponding exponentially smoothed estimate when σ^2 is known. The results of this comparison are displayed in Table XV. For purposes of comparison the initial guess of mean demand was set at zero for both estimation techniques, that is, $S_o(Y) = 0$ and $\theta_o = 0$, and random demand generated with θ = 100 and σ = 10. The mean demand was estimated using each method with various choices of the observed value weighting factors, α in the case of smoothing, and σ_o^2 for the Bayesian method. The entries in the Table reveal that the Bayesian approach yields better estimates than smoothing in the early periods even for a relatively large smoothing constant. A choice of $\sigma_o^2 \geq 2\sigma^2$ resulted in better

TABLE XIV

AVERAGE BAYES ESTIMATE OF THE MEAN
FOR VARIOUS PRIOR PARAMETER VALUES
($\theta = 100$, $\sigma = 10$)

σ_o^2 \ θ_o	0		20		25		33		50		100	
	5	100	5	100	5	100	5	100	5	100	5	100
20	50.2	95.3	60.2	96.2	62.7	96.5	66.9	96.9	75.2	97.7	100.2	100.0
25	55.8	96.2	64.7	97.0	66.9	97.2	70.6	97.5	78.0	98.1	100.2	100.0
33	62.8	97.1	70.3	97.7	72.1	97.9	75.3	98.1	81.5	98.6	100.3	100.0
50	71.7	98.1	77.4	98.5	78.9	98.6	81.3	98.7	86.0	99.1	100.3	100.0
100	83.7	99.1	87.0	99.3	87.8	99.3	89.2	99.4	92.0	99.5	100.3	100.0
200	91.3	99.5	93.1	99.6	93.6	99.7	94.3	99.7	95.8	99.8	100.4	100.0
300	94.1	99.7	95.4	99.8	95.7	99.8	96.2	99.8	97.3	99.9	100.4	100.0
400	95.6	99.8	96.6	99.8	96.8	99.9	97.2	99.9	98.0	99.9	100.4	100.0
500	96.6	99.8	97.3	99.9	97.5	99.9	97.8	99.9	98.5	99.9	100.4	100.0

TABLE XV

AVERAGE ESTIMATE OF MEAN DEMAND
BY BAYES AND SMOOTHING

$\theta = 100$ $\sigma = 10$	$\theta_o = 0$ $S_o(Y) = 0$	Number of Periods				
Estimator	Parameter Values	5	10	15	50	100
BAYES	$\sigma_o^2 = \sigma$	83.7	91.0	93.7	97.9	99.1
	$2\sigma^2$	91.3	95.4	96.8	98.8	99.5
	$3\sigma^2$	94.1	96.9	97.8	99.2	99.7
	$4\sigma^2$	95.6	97.7	98.3	99.3	99.8
	$5\sigma^2$	96.6	98.2	98.7	99.4	99.8
SMOOTHING	$\alpha = 0.1$	41.1	65.1	79.3	99.2	100.4
	0.2	67.3	89.1	96.4	99.8	100.6
	0.3	83.2	96.9	99.5	99.9	100.8
	0.4	92.1	99.0	100.0	99.9	100.9
	0.5	96.5	99.5	100.3	99.9	100.9

estimates by the Bayes technique through 15 periods than the smoothing estimator with the usual smoothing constant 0.20.

To determine the effect of the prior parameters on the sample risk obtained by the forecasting techniques the same model was used to generate demand for 100 periods and the experiment replicated 1000 times. For the case where the prior estimates were the best possible, that is, $S_o(Y) = \theta$ and $\theta_o = \theta$, the results are presented in Table XVI. For

various choices of weighting factors, the desired risk was held constant at 0.05. The results are quite remarkable.

TABLE XVI

SAMPLE RISK OBTAINED BY BAYES
AND SMOOTHING OVER 1000 REPLICATIONS
(GOOD INITIAL VALUES)

$\theta = 100$ $\sigma = 10$	$\theta_o = 100$ $S_o(Y) = 100$	Number of Periods				
Estimator	Parameter Values	5	10	15	50	100
BAYES	$\sigma_o^2 = \sigma^2$	.066	.056	.057	.057	.055
	$2\sigma^2$	.069	.057	.058	.057	.055
	$3\sigma^2$	.072	.058	.058	.057	.055
	$4\sigma^2$	.073	.059	.058	.057	.055
	$5\sigma^2$	.074	.059	.058	.057	.055
SMOOTHING	$\alpha = 0.1$	.060	.051	.057	.057	.063
	0.2	.059	.061	.061	.061	.069
	0.3	.066	.073	.069	.070	.079
	0.4	.073	.077	.079	.076	.083
	0.5	.082	.082	.083	.084	.091

Except for a few cases the Bayes forecasting method provides a sample risk closer to the desired risk than the smoothing process. Also noteworthy is the fact that the Bayes method again appears much more stable than smoothing. This is indeed the case as indicated by the sample standard deviation

in Table XVII. For only the early periods and for small

TABLE XVII

SAMPLE STANDARD DEVIATION OF BAYES AND SMOOTHING ESTIMATORS OF MEAN DEMAND

$\theta = 100$ $\sigma = 10$	$\theta_o = 100$ $S_o(Y) = 100$	Number of Periods				
Estimator	Parameter Values	5	10	15	50	100
BAYES	$\sigma_o^2 = \sigma^2$	3.95	3.26	2.84	1.95	1.69
	$2\sigma^2$	4.25	3.40	2.94	1.97	1.72
	$3\sigma^2$	4.36	3.41	2.94	1.96	1.73
	$4\sigma^2$	4.46	3.47	2.95	1.95	1.72
	$5\sigma^2$	4.48	3.48	2.97	1.95	1.72
SMOOTHING	$\alpha = 0.1$	2.31	2.64	2.68	2.61	2.73
	0.2	3.41	3.68	3.63	3.60	3.65
	0.3	4.30	4.51	4.40	4.49	4.44
	0.4	5.10	5.27	5.17	5.30	5.16
	0.5	5.96	6.03	5.90	6.06	5.88

values of the smoothing constant does the smoothed estimate demonstrate a smaller variance than the corresponding Bayesian estimate. As more and more observations are made, the sample standard deviation of the Bayes estimate becomes smaller while the variance of the smoothed estimate remains about the same.

The same experiment was repeated with $\theta_o = 0$ and $S_o(Y) = 0$ to determine how quickly the two forecasting methods could demonstrate sensitivity to the actual data and obtain a sample risk close to the desired risk 0.05. In this case also the experiment was replicated 1000 times in order to obtain a degree of significance in the results. The sample risks obtained by both methods are displayed in Table XVIII. The

TABLE XVIII

SAMPLE RISK OBTAINED BY BAYES

AND SMOOTHING OVER 1000 REPLICATIONS

(POOR INITIAL VALUES)

$\theta = 100$ $\sigma = 10$	$\theta_o = 0$ $S_o(Y) = 0$	Number of Periods				
Estimator	Parameter Values	5	10	15	50	100
BAYES	$\sigma_o^2 = \sigma^2$	.635	.294	.187	.079	.071
	$2\sigma^2$	.310	.159	.107	.068	.062
	$3\sigma^2$	.211	.128	.090	.066	.061
	$4\sigma^2$	.168	.109	.084	.065	.061
	$5\sigma^2$	.146	.097	.080	.065	.060
SMOOTHING	$\alpha = 0.1$	1.000	.992	.737	.067	.063
	0.2	.995	.390	.145	.061	.069
	0.3	.761	.137	.080	.070	.079
	0.4	.365	.088	.080	.076	.083
	0.5	.187	.084	.083	.084	.091

sample standard deviations of the estimators did not change appreciably from the previous case and are not presented. The results are as expected. When a small value is chosen for the smoothing constant the long term results are fairly accurate, but the results in the early periods are far from satisfactory. If a larger value is chosen for α the smoothed estimate adjusts much more rapidly but gives poorer results in the long run due to the increased variance. The Bayes estimator adjusts quite rapidly for $\sigma_o^2 = 5\sigma^2$ and the long term results for all values tested are fairly accurate.

VI. CONCLUSIONS AND RECOMMENDATIONS

Before summarizing the conclusions reached in this study and the recommendations which follow, it may be worthwhile to present briefly a few of the conclusions reached by Zehna and Ornek in similar studies [Refs. 10, 11, and 6].

1. It is difficult to judge exponential smoothing on a theoretical basis due to a lack of knowledge concerning the probability distributions of the smoothing estimators.

2. Many of the analytical results obtained by Brown [Ref. 2] are asymptotically valid only although applied to the finite demand case.

3. The source of much of the difficulty in determining the distributions of the smoothing estimators is in the use of MAD as an estimate of variability.

4. On the basis of simulation results the simplified MLE approach is to be preferred in those cases in which smoothing is presently employed by NavSup.

A. CONCLUSIONS

It has been pointed out numerous times throughout this thesis that the numerical results and comparisons made are based on computer simulation sample outcomes. However, the experimentation was conducted with great care taken in the handling of random numbers and sample size to assure a degree of significance in the results. The results of the various comparisons of MLE and exponential smoothing are quite consistent and reinforce the conclusions reached in the above

references. Without regressing to the specific numerical examples presented previously in this thesis it can be concluded that on the basis of the analysis performed in this study the smoothing estimators are more variable than any form of maximum likelihood estimation and prediction method tested in the linear or constant mean models. The source of most of the variability in the smoothed estimates appears to be in the MAD estimate of the variance in the demand distribution. This results in far less consistent risks obtained by smoothing than the comparable risks obtained using MLE. As for the ease or difficulty of computation using the various methods, the exact MLE computations are certainly more involved but probably not to the extent of resulting in significantly increased computation time. In any case the trade-off between computational accuracy and computation time is a managerial decision. In the constant mean model the simplified MLE technique is certainly not more difficult to handle computationally than smoothing and has been shown to produce significantly more stable and accurate results. The Bayesian approach, while not immediately applicable without further study, appears to offer a procedure which, although subjective in nature, has a sounder theoretical basis than does smoothing. On the basis of simulation the Bayes method has been shown to produce good estimates, be relatively flexible and sensitive, and demonstrate less variability than exponential smoothing.

B. RECOMMENDATIONS

For the reasons discussed above it is strongly recommended that NavSup or other agencies using smoothing for demand forecasting under the assumptions of the models discussed in this thesis give serious consideration to at least testing other forecasting techniques under a variety of circumstances to compare and contrast the results. This testing should involve the use of actual historical demand data so the results obtained by the alternate forecasting techniques can be compared in retrospect with the forecasting results which were obtained by means of exponential smoothing. Because of its intuitive appeal and computational simplicity exponential smoothing should be investigated further with prime interest in discovering a method to replace MAD as an estimate of variance. If this can be accomplished it may turn out that smoothing is in some respects an optimal method of forecasting. It is also recommended that further study be made of the Bayesian approach developed in this thesis to determine its applicability in actual situations. It would be rewarding to develop a Bayesian procedure which could be used to jointly estimate the mean and variance of the demand distribution and yet be relatively simple to apply. In cases such as those involving a constant proportionality between the mean and variance of the underlying distribution it is speculated that an ad hoc procedure could be developed that would involve an "a priori" estimate of the variance and then, after data are observed, the use of the posterior distribution to estimate the variance as a function of the posterior mean.

Finally it is recommended that an investigation be made into the use of smoothing, MLE, Bayesian methods, or other forecasting approaches in the case of a time dependent mean when demand is normally distributed but mean demand is subject to periodic increases or decreases. In such cases the estimating procedure must not only be able to yield accurate estimates but also be able to detect the changes in the mean of the underlying demand distribution and adjust the forecasts accordingly [Refs. 3, 7, and 9].

APPENDIX A

IDENTITIES USED IN MLE COMPUTATIONS

As discussed in Section IV, one of the disadvantages often attributed to maximum likelihood estimation techniques applied to demand forecasting is the necessity of storing the historical account of demand data for the entire period of interest. The following results show that the above assertion is not valid and that only certain totals are required to be stored much the same as with exponential smoothing.

A. CONSTANT MEAN MODEL

$$\hat{a} = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{a})^2}{n - 1}}$$

but $\sum_{i=1}^{n} (y_i - \hat{a})^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - 2\bar{y} \sum_{i=1}^{n} y_i + n\,\bar{y}^2$

$$= \sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2$$

so that $S_y = \sqrt{\dfrac{\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2}{n - 1}}$

which shows that $\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n} y_i^2$ are the only data that must be retained from period to period.

B. LINEAR MODEL - Y-INTERCEPT = 0

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

$$s_{y \cdot x} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{b} x_i)^2}{n - 1}}$$

but $\sum_{i=1}^{n} (y_i - \hat{b} x_i)^2 = \sum_{i=1}^{n} y_i^2 - 2\hat{b} \sum_{i=1}^{n} x_i y_i + \hat{b}^2 \sum_{i=1}^{n} x_i^2$

$$= \sum_{i=1}^{n} y_i^2 - \hat{b} \sum_{i=1}^{n} x_i y_i$$

so that

$$s_{y \cdot x} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - \hat{b} \sum_{i=1}^{n} x_i y_i}{n - 1}}$$

and the only additional information which must be stored is

$\sum_{i=1}^{n} y_i^2$.

C. GENERAL LINEAR CASE

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_1^2 - n\bar{x}^2} = \frac{N}{D}$$

$$\hat{a} = \bar{y} - \bar{x}\hat{b}$$

$$S_{y \cdot x} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i)^2}{n - 2}}$$

Now omitting the limits on the sums for simplicity

$$\sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i)^2 = \sum y_i^2 - 2\hat{a}\sum y_i - 2\hat{b}\sum x_i y_i$$

$$+ \hat{b}^2 \sum x_i^2 + 2\hat{a}\hat{b}\sum x_i + n\,\hat{a}^2$$

which becomes after substituting for $\hat{a}$ and combining terms

$$\sum y_i^2 - n\,\bar{y}^2 - 2\hat{b}(\sum x_i y_i - n\,\bar{x}\bar{y}) + \hat{b}^2(\sum x_i^2 - n\,\bar{x}^2) \ .$$

But the quantities in parenthesis are just N and D respectively so noting that

$$\hat{b}^2 \cdot D = \hat{b} \cdot N$$

it follows that

$$\sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i)^2 = \sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2 - \hat{b}N$$

and

$$S_{y \cdot x} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2 - \hat{b}N}{n - 2}}$$

Note that D depends only upon the number of periods n and if the demand is being followed over successive periods of time the x_i's are just the first n integers so that the value of D can be stored for each period of time only and not calculated for each particular item of supply.

LIST OF REFERENCES

1. Brown, R. G. and Meyer, R. F., "The Fundamental Theorem of Exponential Smoothing," Operations Research, vol. 9, No. 5, September-October 1961, pp. 673-685.

2. Brown, R. G., Smoothing, Forecasting, and Prediction of Discrete Time Series, Prentice Hall, 1963.

3. Chernoff, H. and Zacks, S., "Estimating the Current Mean of a Normal Distribution which is Subjected to Changes in Time," The Annals of Mathematical Statistics, vol. 35, No. 3, September 1964, pp. 999-1018.

4. Ferguson, Thomas S., Mathematical Statistics, A Decision Theoretic Approach, Academic Press, 1967.

5. Hald, A., Statistical Theory with Engineering Applications, John Wiley and Sons, 1952.

6. Ornek, O., An Analysis of Mean Absolute Deviation Variability, M.S. Thesis, Naval Postgraduate School, October 1969.

7. Packer, A. H., "Simulation and Adaptive Forecasting As Applied to Inventory Control," Operations Research, vol. 15, No. 4, July-August 1967, pp. 660-679.

8. Pratt, John W. and Raiffa, Howard, Introduction to Statistical Decision Theory, McGraw-Hill, 1965.

9. Rao, Ambar G. and Shapiro, Arthur, "Adaptive Smoothing Using Evolutionary Spectra," Management Science, 1970, pp. 208-218.

10. Zehna, Peter W., Some Remarks on Exponential Smoothing, Naval Postgraduate School, Technical Report No. 72, December 1966.

11. ____________. Forecasting Errors Using MAD, Naval Postgraduate School, April 1969.

12. ____________. Probability Distributions and Statistics, Allyn and Bacon, 1970.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Professor Peter W. Zehna, Code 55 Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1
4.	Department of Operations Analysis (Code 55) Naval Postgraduate School Monterey, California 93940	1
5.	CPT John A. Coventry 5302 St. Elmo Avenue Chattanooga, Tennessee 37409	1

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

A. COMPARISON OF DEMAND FORECASTING TECHNIQUES

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Master's Thesis; March 1971

5. AUTHOR(S) (First name, middle initial, last name)

John Ainsworth Coventry
Captain, United States Army, MSC

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1971	77	12

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

For periodic review inventory models with stochastic demand, the idea of stock-out risk is defined, from which the importance of accurate prediction of demand is deduced. Methods of demand model parameter estimation are investigated and several methods compared on the basis of theoretical soundness, ease of application, and accuracy of estimates based upon the results of extensive computer simulation. The theoretical development of maximum likelihood and exponential smoothing estimators as applied to prediction is presented along with the development of a new Bayesian approach to the problem of demand forecasting.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
prediction of demand						
forecasting of demand						
exponential smoothing						
applied Bayesian estimation						
maximum likelihood applications						